

National Reading Conference Policy Brief  
High Stakes Testing and Reading Assessment

Peter Afflerbach  
University of Maryland

September 2004

National Reading Conference Policy Brief  
High Stakes Testing and Reading Assessment

**Executive Summary** 

This National Reading Conference Policy Brief provides information related to high stakes reading tests and reading assessment. High stakes reading tests are those with highly consequential outcomes for students, teachers and schools. These outcomes may include student promotion or retention, student placement in reading groups, school funding decisions, labeling of schools as successful or failing and the degree of community support for a school. The Policy Brief focuses on the popularity of high stakes tests, the uses and misuses of high stakes tests and the consequences of high stakes testing. While high stakes tests are believed by many to be central to efforts to raise school accountability and student achievement, these tests are accompanied by numerous liabilities. These include the following:

- High stakes tests are used with increasing frequency in spite of the fact that there is no research that links increased testing with increased reading achievement.
- High stakes tests are limited in their ability to describe students' reading achievement.
- High stakes tests may be harmful to students' self-esteem and motivation.
- High stakes tests confine and constrict reading curriculum.
- High stakes tests alienate teachers.
- High stakes tests disrupt high quality teaching and learning.

- High stakes tests demand significant allocation of time and money that could be otherwise used to increase reading achievement.
- High stakes tests are used with increasing frequency to characterize and label young children who are in early developmental stages of reading.
- High stakes tests most often come with caveats related to the accuracy of scores and suitability of uses of scores, and these caveats are widely ignored.

### **Introduction**

We begin this Policy Brief with  the critical assumptions. First, reading assessment is a useful tool in the service of improving the teaching and learning of reading. This focus for reading assessment allows us to assess and understand the strengths and needs of each of our students so we may teach them well. Second, all reading assessment must be clearly and carefully tied to an informed understanding of what reading "is." We are fortunate to have a rich knowledge base that describes the nature of reading and its development (Clay, 1979; Heath, 1983; Snow, 2002). From this research we are able to construct an understanding of reading that should guide our efforts to design and use reading assessment effectively. Third, reading assessment must reflect our most current knowledge in the science of assessment. The historical progress in our understanding of reading is paralleled by enhanced understanding about how to develop and use effective assessments (Pellegrino, Chudowsky & Glaser, 2001). Advances in theoretical and practical knowledge require that high quality reading assessments correspond to current knowledge of the content and processes that should be assessed in reading (e.g., student growth and achievement). In addition, better assessments equate with improved materials, procedures and contexts used in conducting

the assessment, and should reinforce the inferences we make about students' reading development and achievement from assessment data. Never before has there been such a rich array of means for assessing students' reading development and achievement. Yet, high stakes testing may repress the realization of high quality reading instruction and assessment and it poses a threat to the development of students who are accomplished, lifelong readers.

*Popular beliefs related to high stakes tests*

The public, in general, supports high stakes testing (Afflerbach, 2002). Why do tests enjoy such popularity? Many people believe that high stakes tests are fair. Under typical standardized testing conditions, it is assumed that no student receives preferential treatment because each student has the same interaction and degree of support from the person administering the test. But examination of the assumption of fairness raises several questions. Will all students who are tested share the same amounts and types of prior knowledge and experience? Students who lack particular world knowledge and testing experience will not have experiences similar to those students who possess this knowledge and experience. Students who are not familiar with high stakes testing culture will be at a disadvantage. Will anxiety influence students' test results? Certain students become overly anxious when they are tested and this influences test performance. Thus, the inferences we make about students' reading achievement from a single score on a test that is influenced by students' prior knowledge, experiences and level of anxiety must be made with caution.

A second reason for test popularity is that many people believe high stakes tests are scientific. The vast majority of commercial and statewide reading tests are the result

of considerable time and effort invested in developing and piloting the tests. Through adherence to what are for most people abstract notions of validity and reliability, tests can create a "scientific" aura. Tests have the ability to reduce and summarize complexities of reading to single raw scores and percentile rankings, and in doing so they appear almost magical (Lemann, 1999). Many members of school communities are in the habit of receiving test scores and taking them at face value: this serves to maintain their use and considerable reputation. Yet, if we examine what tests tell us about student reading and compare that with what we hope for all our students, we will find a considerable mismatch. The very high stakes tests that are believed to be scientific are actually severely limited in their ability to describe the wide range of reading achievement that most states and school districts set forth in their formal learning goals and standards statements (Davis, 1998). Understanding short texts and answering questions about them, required of all students taking high stakes tests, is but a small slice of what we expect accomplished student readers to do. In this sense, high stakes tests are an exceedingly thin measure of reading achievement and reading ability.

Finally, high stakes tests are popular because they are familiar. These tests are administered with increasing frequency, as No Child Left Behind demands their use on an annual basis from grades 3 through 8. There are few adults who have not experienced first-hand the culture of high stakes testing in schools. Throughout our school years, high stakes tests have been used to make important decisions about admissions, awards and access to opportunities. Testing appears natural, given its tradition. Yet, we are in the midst of ongoing changes in assessment (Pellegrino, Chudowsky & Glaser, 2001). We know that the current array of high stakes reading tests represents views of both reading

and the assessment of reading that are decades old. The format of current high stakes reading tests limits our ability to know how students read critically, how they evaluate what they read and how they use the knowledge that they gain through reading (Afflerbach, 2002).

*Concerns with high stakes testing*

Against this background of the American public's general acceptance of high stakes testing practice in reading, we offer the following warnings. Each is intended to encourage reflection on the high stakes testing regimen that is firmly entrenched in schools across the United States. Examples and evidence that suggest why high stakes testing practices should be challenged and changed follow each of the bulleted statements.

- **High stakes tests are used with increasing frequency in spite of the fact that there is no research that links increased testing with increased reading achievement.** It is ironic that frequent high stakes testing in reading accompanies federal and state initiatives to certify that all reading instruction is based on the results of scientific research. In reality, no research has been conducted that demonstrates a cause and effect relationship between increased high stakes testing and improvement in reading achievement scores. The massive testing of reading that is now federally mandated in grades 3 through 8 results in valuable class time taken from the instruction of reading and given to test preparation and administration. Valuable funds are taken from the teaching of reading and put into the testing of reading. In effect, the reduction in time for classroom reading

- instruction that is caused by high stakes tests may well reduce levels of student reading achievement.
- **High stakes tests are limited in their ability to describe students' reading achievement.** A high stakes test score represents a single sample of what a student reader does on a standardized test. This score is not at all representative of the full range of reading and using what is read that marks accomplished student readers. The single score provides only a snapshot of student performance. Classroom teachers may use high stakes reading test scores as one piece in an array of evidence that suggests certain levels of student achievement and possible instructional strategies, but parents and administrators often focus exclusively on the single score. In the decision-making processes of a talented teacher, determining a convergence of evidence that different reading assessments provide and identifying trends in that evidence are critical abilities. In no situation should a single high stakes reading test score be used to make important educational decisions. High stakes tests may well er represent the accomplishments of students and their teachers because these tests have a severely limited ability to describe complex reading and reading-related performances that mark the accomplished teaching and learning of reading.
  - **High stakes tests may do more harm than good.** One of the greatest challenges related to high stakes reading tests is to fully anticipate their consequences. Some students always do well on tests. In contrast, there are students who never do well on tests. These students may be all too familiar with the routine of their reading failure and public disclosure of their failures, as they struggle through hours of

high stakes testing. When tests are norm-referenced, half of the students taking the test may be determined to be “below average.” A less able student's improvement in reading achievement, while substantial, may be reflected in a large change in percentile ranking but no change in the "below average" label. Percentile rankings and raw scores are assigned to individual students, creating communities of "haves" and "have-nots" when it comes to reading achievement. Labels for students who do not do well on tests are accompanied with a myriad of related consequences. These include lowered expectations, differential treatment in the classroom and decreased perseverance for those labeled as low-achieving readers.

- **High stakes tests confine and constrict reading curriculum.**  classrooms everywhere, there is tremendous pressure to improve test scores. Reading curriculum may be developed or chosen on the sole merit of being closely related to the content and form of the high stakes reading test. This is misguided: there is no current high stakes test of reading that approaches a representation of reading and reading development that is informed by our current understanding of reading. For example, we have no high stakes tests that assess the reading that takes place in online environments and few that provide rich measure of critical reading. When reading curricula are developed solely in relation to high stakes tests or when they are molded to “fit” a single reading assessment, the breadth and positive outcomes any curriculum is diminished. Tests contribute to the constriction, or shrinking of reading curricula (Frederiksen, 1984). Were the majority of standardized high stakes reading tests worth teaching to, this would

- not be an issue. The point is that most high stakes tests represent an over simplistic view of reading and have anarrow focus on particular reading skills and strategies.
- **High stakes tests alienate teachers.** Teachers are often conflicted about their stances towards testing and their students (Calkins, Montgomery, Santman & Falk, 1998). What of the teacher who does not endorse high stakes testing yet knows that her students must be adequately prepared to take such tests? What of the teacher who has worked for years to develop an array of engaging reading instruction lessons and is forced to give up this good work? When testing concerns override teacher professionalism, curriculum decisions may be made according to how well reading instructional materials mirror a test format, and not according to accomplished teachers' knowledge. Each day spent teaching to a test that narrowly samples from what reading is can contribute to decreased teacher motivation and enthusiasm (Smith, 1991). A high stakes test that is administered in September or October with scores reported back in April or May is of little or no instructional use. Test scores that are months old are silent on the matter of progress in reading that a student may make during the interim of taking the test and getting scores back. This particular lack of suitability of high stakes test results, when teachers may have strong and useful assessment alternatives, is troublesome.
  - **High stakes tests disrupt high quality teaching and learning.** In addition to the negative consequence of high stakes tests constricting the reading curriculum, they may also take considerable time away from the teaching of reading. For

every school hour spent on reading test preparation and administration, an hour is taken from the instructional day. Time given for practice to take tests is time that could be spent on pursuing diverse instructional goals related to reading. Recent mandates at the national and state and district levels have resulted in an array of non-trivial forced changes to reading instruction. New regulations require that significant blocks of time be dedicated to test preparation exercises for all students. Teachers who have a long tradition of accountability to their students (and to themselves as professionals) find the imposition of test preparation, in the name of accountability, to be in opposition to what the teachers believe is best for their students.

- **High stakes tests demand significant allocations of time and money that could be otherwise used to increase reading achievement.** When schools devote time and effort to testing and preparing students to take tests, they spend considerable sums of money to do so. This money might be spent in other ways, including enriched reading curriculum and teachers' professional development related to assessment (Stiggins, 2002). High stakes tests are expensive to purchase, prepare for, administer and score. Initiatives to support testing take from other, worthy initiatives related to fostering students' reading development.
- **High stakes tests are used with increasing frequency to characterize and label young children who are in early developmental stages of reading.** Tests must be used appropriately. Children's growth and experiences related to reading vary widely prior to formal schooling, as does their experience with formal testing situations. We expect this varied experience to influence the skills, strategies,

motivation and conceptions of reading that young children bring to school. In contrast to this variety of experiences and abilities, high stakes tests force the labeling of young children and assignment to differential instruction that might not be appropriate or effective. Additionally, it is doubtful that most young children have extensive standardized testing experience, so the very act of placing these children in such a situation introduces factors of familiarity and anxiety as possible influences on test performance (National Association for the Education of Young Children, 2003).

- **High stakes tests most often come with caveats related to the accuracy of scores they produce and the suitability of uses of scores, and these caveats are widely ignored.** Commercially produced reading tests and those created for statewide and federal high stakes decision making regularly feature strong guidance related to the appropriate uses and misuses of test scores (Hamilton, Stecher, & Klein, 2002). Among the most frequent caveats is the admonition not to use a single high stakes reading test score to make educational decisions. This caveat is based on the understanding that single test scores represent only one measure of student readers and that test scores are subject to natural variation and sampling error, as with the political polls that are conducted regularly in the United States. This means that a student's single high stakes reading test score falls within a range of scores that represent the students' actual achievement level. When high stakes decisions are made using such unstable scores, the decisions may be faulty and costly. Psychometricians and test developers are fully aware of the dangers of such practice, yet it continues unabated.

## Recommendations for improving reading assessment

Given the limitations of high stakes test scores for describing the breadth and depth of student reading achievement and given the history of inappropriate use and negative consequences related to high stakes test, effective reading assessment will have several characteristics.

- **Reading assessment should reflect performance over multiple time points with various texts and purposes.** We must provide students with opportunities to demonstrate their reading growth and achievement in situations that reflect their daily lives as readers. This means that assessment is conducted to gather both formative and summative information, to describe the detail of learning and achievement of student reading and to help teachers determine what reading instruction and experience is best suited to each student.
- **Assessment should measure a wide range of skills with a variety of formats and responses.** Our understandings of how students construct meaning from text and how to best assess this process continually improve. We have the means to use diverse reading assessments, including teacher questioning, performance assessments, portfolios and high stakes tests to fully describe students' accomplishments. Inferences about student reading ability, teacher accountability and school goodness are often made from the single picture that a high stakes test score provides. But there are potentially dire consequences if this single view of the reader is not augmented by other, more regular, reading assessment information.

- **Assessments should follow ethical guidelines of the American Educational Research Association, standard practices of the American Psychological Association and recommended practices of the International Reading Association.** The national and international organizations that represent researchers and practitioners in reading and their state of the art knowledge have clear guidelines for effective and ethical assessment practices. Each set of guidelines reflects the dual focus of working in the students' best interests while attempting to gather the most useful assessment information (American Educational Research Association, 2000; American Psychological Association, 2001; International Reading Association, 1999).
- **Assessments should provide clear distinctions between the acquisition of reading skills and the effective use of the skills for various purposes.** We have voluminous evidence that reading is a developmental in nature, and we should expect that reading assessments reflect this fact. An assessment that describes students' ability to decode single syllable, phonetically regular words is important for early readers, as is assessment that describes more advanced readers' ability to critically interpret and evaluate persuasive writing.
- **Assessments should provide students with useful information about their developmental accomplishments with clear suggestions for improvement.** Students should not be outsiders to the culture of reading assessment (Black & William, 1998). Assessment that is formative can help shape student readers' development. In addition, as student readers become familiar with the nature of

assessment, they may assume responsibility for assessing themselves—a hallmark of the independent and successful reader.

- **Assessments should provide teachers with useful diagnostic information that can be linked to classroom instruction.** Effective reading instruction is fueled by an array of formative assessments that teachers use to create, revise or maintain their reading instruction. The teachable moment in reading revolves around the professional teacher knowing students and knowing their craft, and this knowledge is informed by regular, diagnostic reading assessment. In contrast, high stakes tests provide summative information that is months old by the time it is reported back to schools and teachers.
- **Assessments should provide parents with comprehensible explanations of their children's progress and achievement with suggestions for enhancing their involvement with their children's literacy development.** The communication of quality reading assessment information to parents increases the possibilities of establishing strong home-school relationships. Reading assessment information must be clearly stated and represented so that parents understand the communication and so that they may act on the information provided to help foster their children's continued reading development.
- **Assessments should provide administrators with data related to specific criteria and standards of performance in order to assess annual progress.** Building and district administrators need reading assessments that describe student and school accomplishment in relation to district and state reading standards. The advent of No Child Left Behind and the mandated high stakes

testing of students in reading in grades three through eight demands that test scores be generated and used to describe student achievement and school accountability.

- **Assessments should be aligned with classroom curricula and instruction.** As we evaluate the suitability of reading assessment, we must focus on how well an assessment or an array of assessments can describe the richness of reading development and achievement that is attributable to high quality reading programs. Assessment that is not aligned with curriculum and instruction may well misrepresent the breadth and depth of student learning and teacher accomplishment.
- **Assessments and testing procedures should be reviewed and revised by school boards, teachers, and parents on a regular basis.** Stakeholders should regularly assess the assessment to determine if it is valid (aligned with curriculum and the construct of reading), up to date (informed by the most recent research on reading and assessment) and useful (providing good information for the varied audiences of reading assessment).

Author Note:

I would like to acknowledge the valuable suggestions and comments provided me by the following officers of the National Reading Conference: Scott Paris, Elizabeth Moje and Deborah Dillon.

### **Suggested Readings**

Afflerbach, P. (2002). The road to folly and redemption: Perspectives on the legitimacy of high-stakes testing. *Reading Research Quarterly*, 37, 348-360.

American Educational Research Association. (2000). *High-stakes testing in preK-12 education*. Retrieved from: <http://www.aera.net/about/policy/stakes.html>

American Psychological Association. (2001). *Appropriate use of high-stakes testing in our nation's schools*. Retrieved from <http://www.apa.org/pubinfo/testing.html>

Black, P., & William, D. (1998). Inside the black box. *Phi Delta Kappan*, 79, 139-148.

Calkins, L., Montgomery, K., Santman, D., with Falk, B. (1998). *A Teacher's Guide to Standardized Reading Tests: Knowledge Is Power*. Portsmouth, NH: Heinemann.

Clay, M. (1979). *Reading: The patterning of complex behaviour* (2<sup>nd</sup> ed.). Portsmouth, NH: Heinemann.

Davis, A. (1998). *The limits of educational assessment*. Oxford, UK: Blackwell Publishers.

Donahue, P., Finnegan, R., Lutkus, A., Allen, N., & Campbell, J. (2001). *The nation's report card: Fourth-grade Reading 2000*. Washington, DC: United States Department of Education.

Frederiksen, N. (1984). The real test bias: Influences of testing on teaching and learning. *American Psychologist*, 39, 193-202.

Hamilton, L, Stecher, B., & Klein, S. (2002). *Making sense of test-based accountability in education*. Washington, DC: Rand.

Heath, S. (1983). *Ways with words: Language, life, and work in communities and classrooms*. New York: Cambridge University Press.

International Reading Association. (1999). *High-stakes assessments in reading*. Retrieved from: [http://www.reading.org/positions/high\\_stakes.html](http://www.reading.org/positions/high_stakes.html)

Lemann, N. (1999). *The big test: The secret history of the American meritocracy*. New York: Farrar, Straus & Giroux.

National Association for the Education of Young Children. (2003). *Early childhood curriculum, child assessment and program evaluation: Building an accountable and effective system for children birth through age eight*. Retrieved from: [http://www.naeyc.org/resources/position\\_statements/positions\\_intro.asp](http://www.naeyc.org/resources/position_statements/positions_intro.asp)

Pellegrino, J., Chudowsky, N., & Glaser, R. (2001). *Knowing what students know: The science and design of educational assessment*. Washington, DC: National Academy Press.

Smith, M. (1991). Put to the test: The effects of external testing on teachers. *Educational Researcher*, 20, 8-11.

Snow, C. (2002). *Reading for understanding: Toward a research and development program in reading comprehension*. Santa Monica, CA: Rand Corporation.

Stiggins, R. (2002). Assessment crisis: The absence of assessment FOR learning. *Phi Delta Kappan*, 83, 758-765.